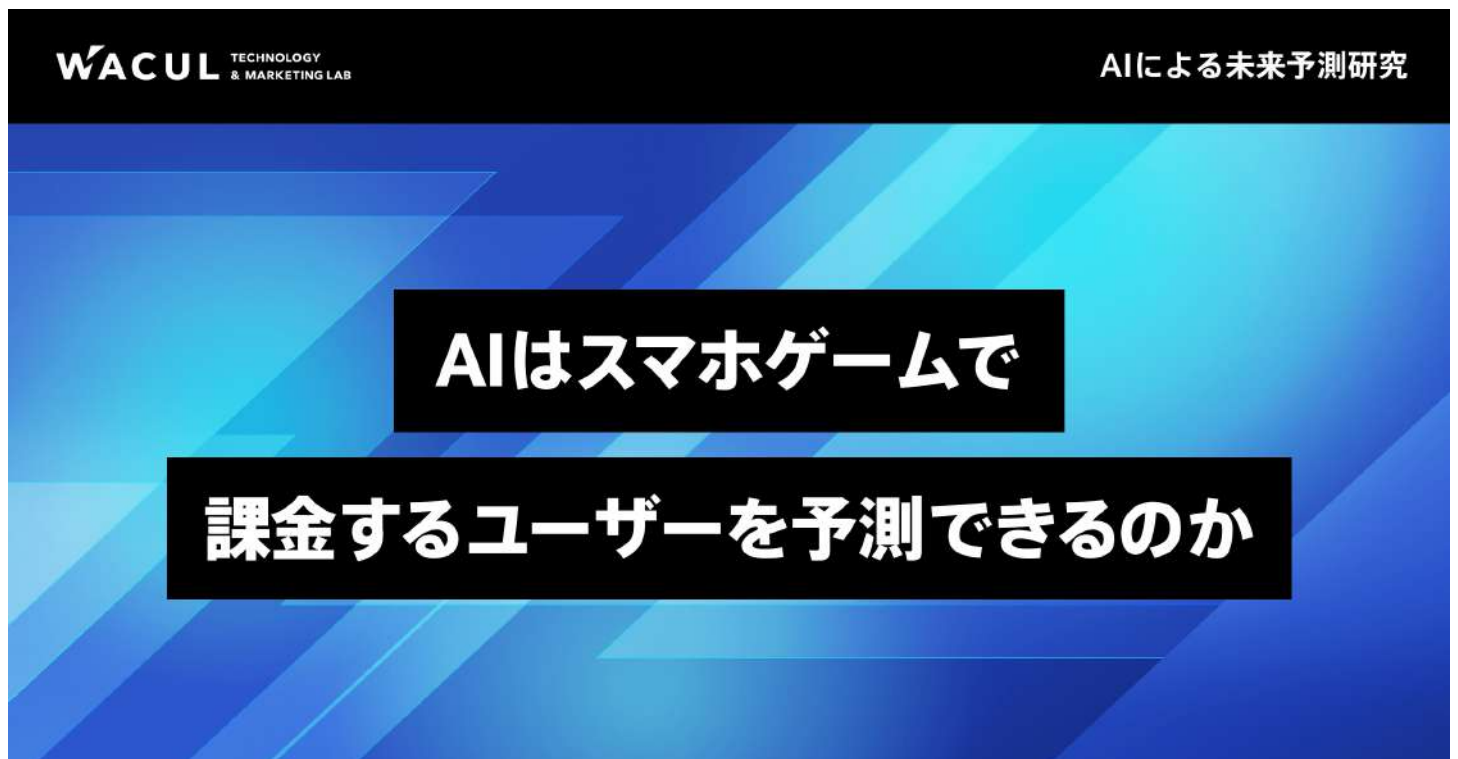


2020.10.08 研究レポート

AIによる未来予測研究～AIはスマホゲームで課金するユーザーを予測できるのか



調査サマリー

- スマホゲームのインストール直後の行動データから、将来課金サービスを利用するかどうかを予測した
- 予測モデルはロジスティック回帰、XGBoost、DNN（ディープラーニング）の3つを用いて精度を比較した
- DNN（ディープラーニング）を用いた場合、ランダム抽出に比べて最大9.03倍の適合率で課金ユーザーを予測できた
- この課金ユーザー予測のアルゴリズムを用いることで、LTVを加味した広告運用のPDCAの高速化や、課金サービスを利用する可能性が高いユーザーに限定したクーポン発行が可能になる

調査の狙い

スマホゲームのビジネスモデル

ゲームアプリが大半を占めるオンラインプラットフォーム市場は、2019年度に1兆2962億円に及び、国内の家庭用ゲーム市場全体の7割以上を占める^{※1}に至っている。

その収益モデルは大きく5つに分類される。

1. フリーミアム(ダウンロード無料・アプリ内課金有)
2. 有料アプリ (ダウンロード有料・アプリ内課金無・売り切り型)
3. ペイドミアム(ダウンロード有料・アプリ内課金有)
4. アプリ内広告(バナー広告、ビデオ広告など広告含有)
5. ダイナミック(ユーザーの利用状況にあわせて収益源を変動するビジネスモデル)

上記のうちアプリで最も多く採用されているビジネスモデルは(1)フリーミアムである^{※2}。

そのためゲーム制作側は、いかに無料ユーザーに課金してもらえらるきっかけを作るかと、いかに課金サービスの利用が見込めるユーザーを獲得するかに苦心している。

※1 10年連続成長の国内ゲーム市場、コロナ禍でさらに需要増加

※2 【App Annie・IDC調査】フリーミアムモデルの課金者の割合は2~5%が最多

コンバージョンを「インストール」に設定する弊害

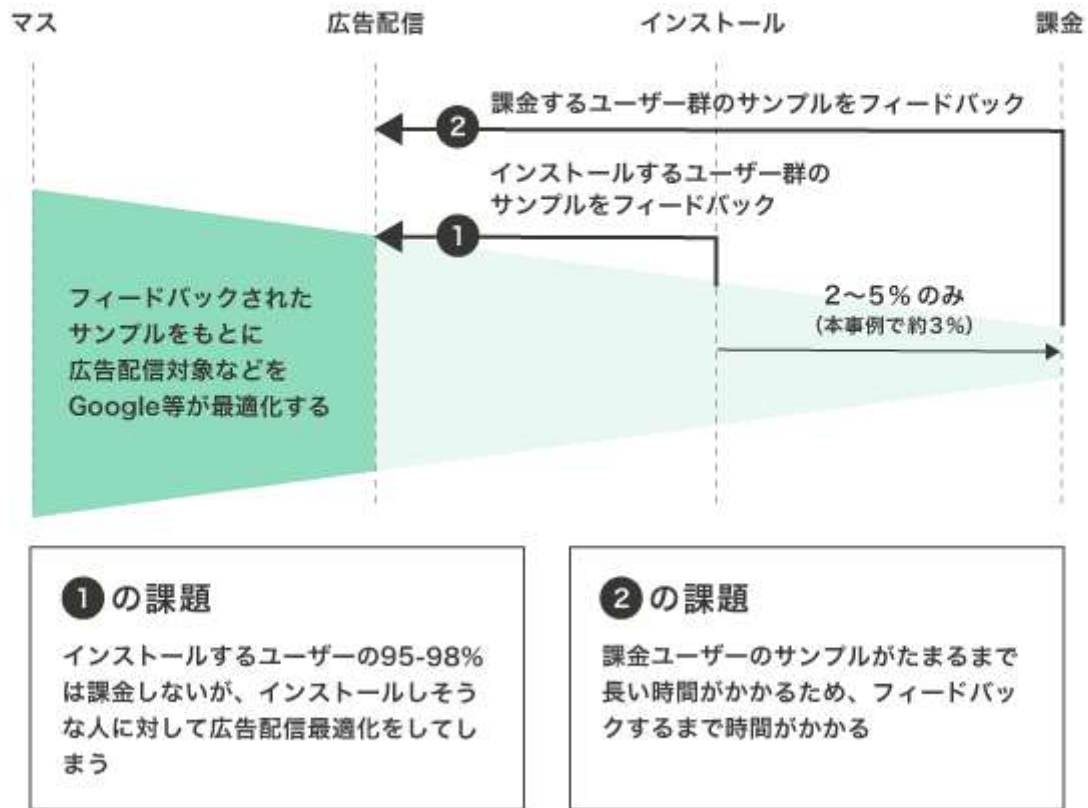
フリーミアムのスマホゲームでは、インストールをしてもすぐやめてしまうユーザーや、無課金でゲームを進めるユーザーが多数いる。そのためコンバージョンを「インストール」にしてPDCAを回しても、課金ユーザーは増えない可能性が高い。特にデジタル広告では、インストール止まりで課金サービスを利用しないユーザーも含めて増やすように最適化されるため、多額の広告費を投下している可能性が高い。

課金ユーザー予測による広告効果改善の可能性

もしコンバージョンを「課金」にすれば、広告の費用対効果は大きく改善する可能性がある。

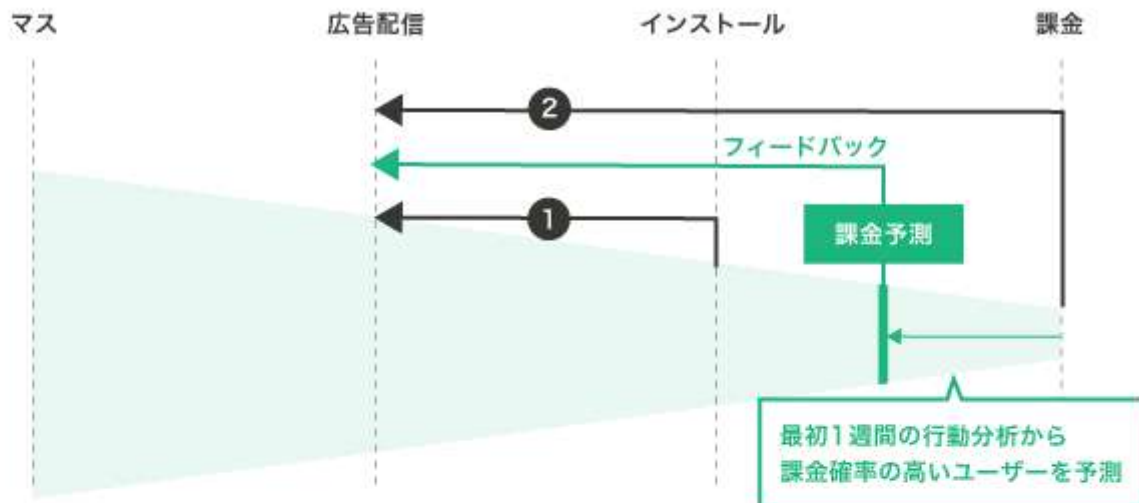
そこで問題となるのが、ユーザーが課金サービスを利用し始めるまでの期間である。ユーザーが実際に課金サービスを利用するまで待っていては時間がかかる。さらに、広告の最適化をかけるために必要な課金ユーザーのサンプル数を集めるにはさらに長い期間が必要となる。そこからデータを分析して広告に反映することを考えると、短期間でPDCAをまわすことは難しい。

図表1: 通常の広告運用のイメージ



その解決策として、インストール後から短期間のユーザー行動から、その後の課金行動を予測し、課金ユーザーを効果的に誘導できている広告を早期に把握、その分析結果を配信広告の運用に活用することで、広告のPDCAを高速化できると考えた。

図表2: 実現したい広告運用のイメージ

**① との比較**

「インストールしたユーザー」に基づくターゲティングではなく、「課金する確率の高いユーザー」に基づくターゲティングができるため、広告の収益性が改善する

② との比較

「課金するユーザー」は長い時間をかけてためる必要があるため、フィードバックが遅くなり、広告のPDCAも遅くなるが、「課金しそうなユーザー」のフィードバックは広告のPDCAのサイクルをもっと早く行える

新たにディープラーニングを用いて「課金しそうなユーザー」をインストール後2週間の行動データから予測してフィードバックすることで、①②の課題を解決する

調査の手法

ゲームアプリの行動データを用いて予測

予測は、行動データから2パターンの分析用データを作成し精度を比較した

- A. ゲーム固有要素に依存しない一般的な特徴量のみでの行動データ（ログイン回数・時刻、課金金額、デバイス種別、IPアドレスから推測された地理情報、ネットワーク回線種別、広告関連情報）
- B. 一般的な特徴量に加え、チュートリアルやイベント消化状況などゲーム固有要素を含んだ行動データ

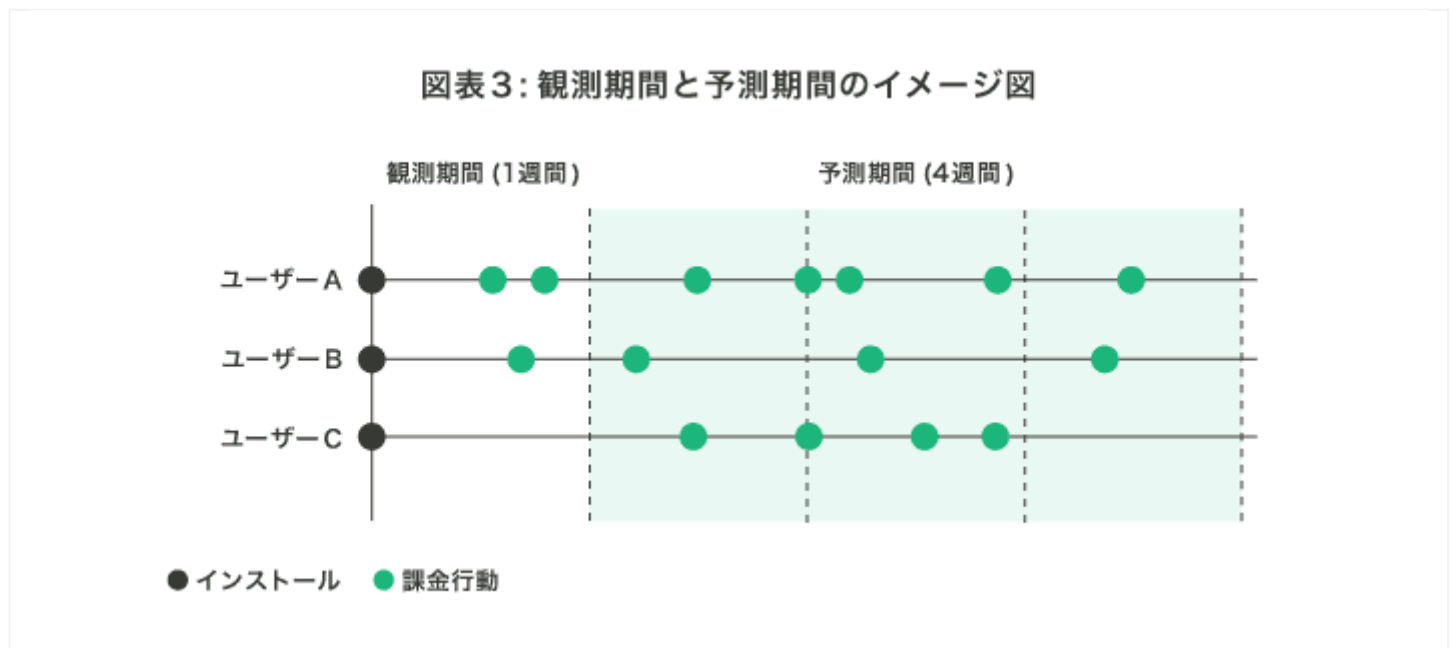
本レポートでは断りがない限り(A)一般的な特徴量のみでの行動データの予測結果を解説する。また、行動データがAndroidとiOSで別テーブルに保存されていたため、別々に予測した。

本分析に関しては、オプトと共同で行った。オプトでは、アプリの広告計測ツールであるadjustやAppsFlyerのシングルソースデータを蓄積できるSpinAppというプロダクトがある。SpinAppでは、どの広告からインストールされたか、インストールされた時刻、その後の起動や各種イベント時刻、端末種別、ネットワーク環境など、ユーザー予測を行うための様々な有益な情報が蓄積できている。

インストール後から1週間の行動データを用いて、その後4週間の間に課金行動を行うかを予測

図3に示すように、インストール後から1週間の観察期間の行動データを用いて、その後4週間の予測期間の間にユーザーが課金行動を一度でも行うかどうか二値分類問題として予測した。

実際は5週間分の行動データを収集し、予測の正しさを評価するために使用した（ユーザー数はAndroidで233,521人、iOSで431,043人。内、約80%を訓練ユーザー、約20%を評価ユーザーとした。以下、実データと呼ぶ）。



分析予測モデル

以下のモデルを用いて課金ユーザー予測と合計課金金額の予測精度を比較した。なお、全ユーザー数に対する課金ユーザー数の割合はAndroidで3.12%、iOSで2.16%の不均衡データであったため、課金ユーザーのオーバーサンプリングを行った。

ロジスティック回帰^{※3}

今回のような二値分類問題で用いられる基本的な分析モデル。計算時間が短く、ベースラインとして用いる。

XGBoost^{※3}

2014年に登場し人気を集めている分析モデル。勾配ブースティングとランダムフォレストを組み合わせた手法で高い汎化能力を誇る。

DNN※3

Deep Neural Network—いわゆるディープラーニング。データ量が多ければ多いほど精度の高い予測が期待できる。今回は4層の隠れ層を持つ多層パーセプトロンを利用した（各隠れ層の次元数は1024、512、256、64次元）。

※3 使用ライブラリ

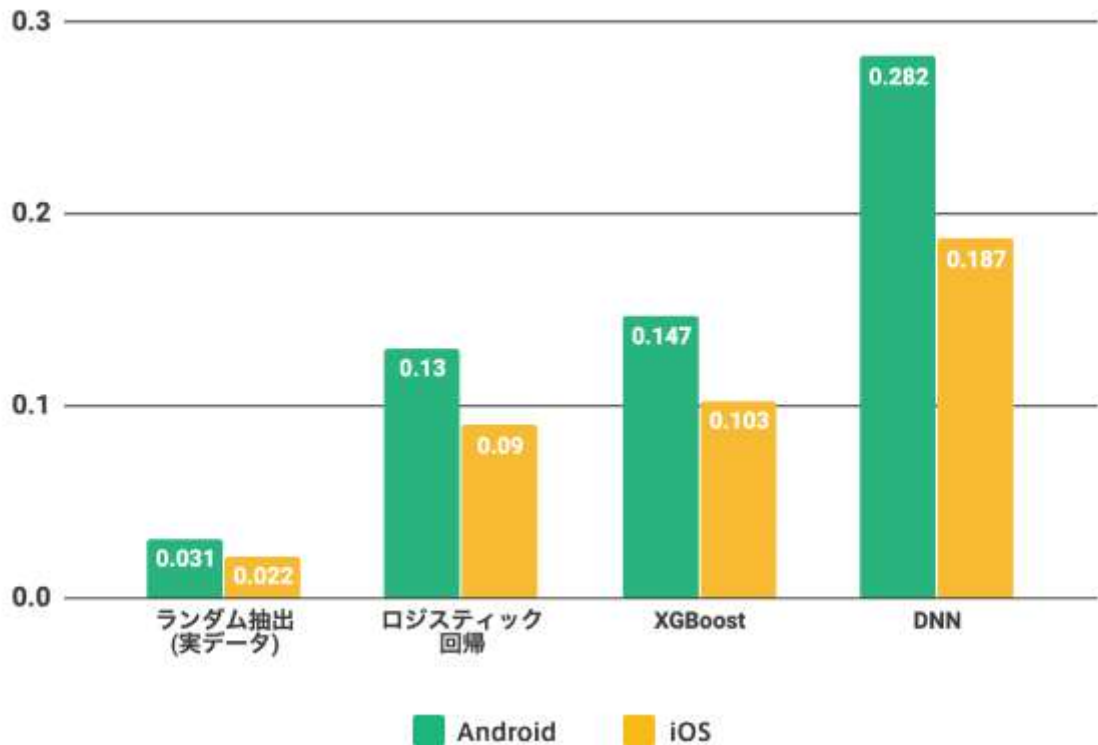
ロジスティック回帰	<code>sklearn.linear_model.LogisticRegression</code>
XGBoost	<code>XGBoost</code>
DNN	<code>TensorFlow</code>

調査結果

ランダム抽出と比べて最大9.03倍の適合率で課金ユーザーを予測

それぞれの予測モデルの適合率（課金サービスを利用すると予測したユーザーが実施に課金サービスを利用した割合。precision、精度ともいう）を比較した（図4）。

図表4：適合率

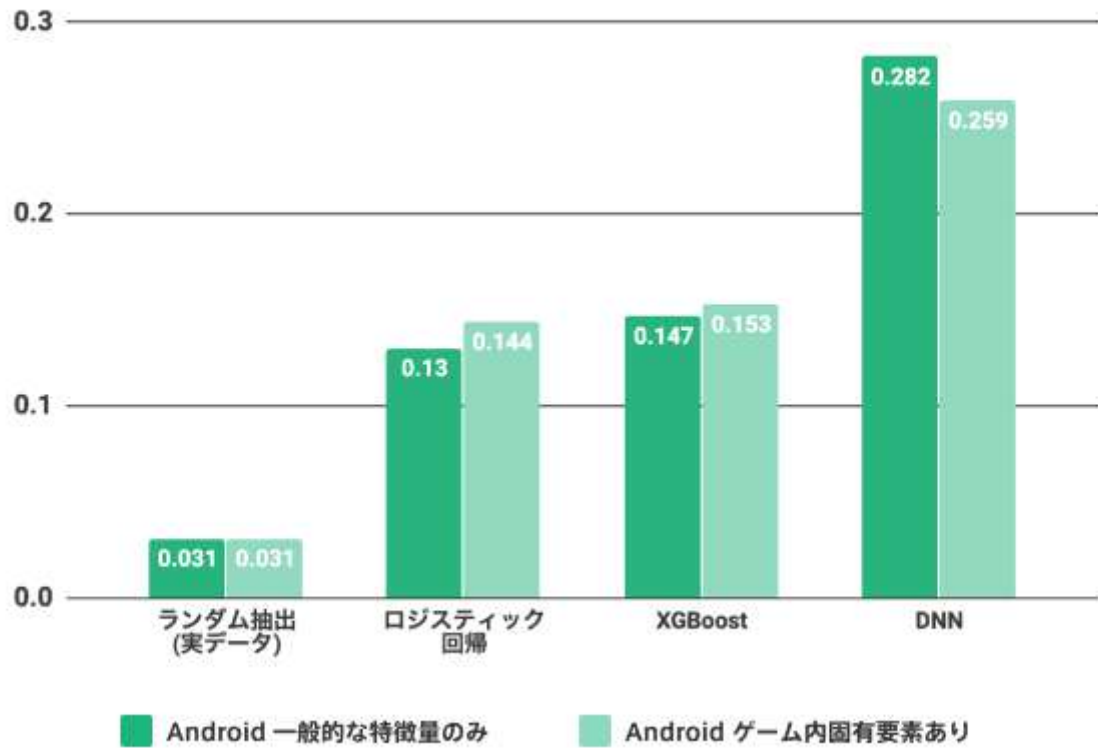


3つのモデルを比較すると、プラットフォームに関わらずDNN（ディープラーニング）が正解率 94 % 程度を維持しつつ適合率が最も高かった。また、ランダム抽出と比較すると最大で9.03倍の適合率で課金ユーザーを特定できた。

ゲーム内のイベントに依らない、汎用性の高い指標で予測可能

同様の予測を(B)ゲーム固有要素を含んだ行動データでも行った。直感的には、チュートリアルやイベント消化状況などのゲーム固有要素によって、課金ユーザーの予測性能が上がるように思われる。しかし実際は、ゲーム固有要素がある場合と比べて、一般的な特徴量のみでの予測の方が、同等かそれ以上の適合率が出ていることがわかった（図5）。

図表5：ゲーム固有要素の有無による適合率の比較



調査の提言

スマホゲームの広告効果改善が期待される

今回の予測データを、実際にGoogle広告にフィードバックする利点は2点ある。

- ブランドサイトやアプリインストールページで離脱したアプリをインストールする可能性が高いユーザーの中から、より課金サービスの利用が見込めるユーザーを獲得するために広告配信の最適化がかけられる
- 類似拡張により、通常のデータのフィードバックでは届かなかった課金サービスの利用が見込めるユーザーへのリーチが期待できる

上記施策を実施した際には改めて報告を行う予定である。

ユーザー毎の合計課金金額の予測も可能

今回は二値分類を実施したが、回帰問題としてユーザー毎の合計課金金額も予測することが可能である。この場合、線形回帰や XGBoost、DNN などの手法を利用できる。

広告配信以外にも、LTV向上のために利用できる

課金ユーザー予測は、広告配信以外にも以下のように利用することでLTVを最大化できる可能性がある。

- 課金サービスの利用が見込めるユーザーにクーポンを発行し課金のきっかけを作る
- 課金サービスの利用が見込めるユーザーをゲーム内イベントで優遇する

工夫によりさらに予測精度が上がるのが期待される

今回は単一の予測モデルを使い、かつ1週間の行動データから将来の課金額を予測したが、以下のような工夫を加えることにより、さらに精度を上げられる可能性がある。

- 1週間よりも長い期間のデータを使い予測する
- DNNとXGBoostをアンサンブルする
- 自動ハイパーパラメータ最適化ツールなどを利用し、モデルを最適化する

コンバージョン後の行動で売上が左右される他業界への応用の可能性

今回の予測からゲーム固有の要素を利用しない一般的な特徴量のみでも一定の予測精度が示されたため、スマホゲーム以外の業界での応用も期待される。

たとえば、人材業界では求人の閲覧数やマイページへのログイン頻度により、その後に転職するかどうかを予測できる可能性がある。EC業界でもは、同様に、閲覧数やログイン頻度を用いてLTVの高い顧客を、また、金融業界では、クレジットカードの新規発行者の利用動向から、メインカードとしてくれるユーザーを予測できる可能性がある。

まずはLTVで振り返る運用から始める

コンバージョン後の行動が重要な業界であっても、LTVを加味して施策を振り返る運用を行っている企業がどれだけあるだろうか。

今回のような予測モデルによる運用改善が技術的に難しい企業であっても、まずはLTVを加味して施策を振り返ることで、予算の使い方や業務フローを見直すことができるのではないだろうか。

付録:各モデルの性能表

Android

	一般的な特徴量のみ			ゲーム固有要素を含めた特徴量		
	ロジスティック 回帰	DNN	XGBoost	ロジスティック 回帰	DNN	XGBoost
正解率 (accuracy)	0.833	0.940	0.850	0.855	0.934	0.856
適合率 (precision)	0.130	0.282	0.147	0.144	0.259	0.153
再現率 (recall)	0.756	0.584	0.781	0.730	0.589	0.784
F値	0.222	0.380	0.247	0.240	0.359	0.255

iOS

	一般的な特徴量のみ			ゲーム固有要素を含めた特徴量		
	ロジスティック 回帰	DNN	XGBoost	ロジスティック 回帰	DNN	XGBoost
正解率 (accuracy)	0.832	0.939	0.850	0.866	0.929	0.855
適合率 (precision)	0.090	0.187	0.103	0.107	0.164	0.107
再現率 (recall)	0.799	0.597	0.824	0.765	0.603	0.832
F値	0.162	0.284	0.183	0.188	0.258	0.189